

EXHIBIT 5

cDNA sequences of two inducible T-cell genes

(unidentified T-cell mediators/cDNA/concanavalin A-inducible)

BYOUNG S. KWON*[†] AND SHERMAN M. WEISSMAN[‡]

*Department of Microbiology and Immunology and Walther Oncology Center, Indiana University School of Medicine, Indianapolis, IN 46223;

[†]Laboratory of Molecular Genetics, Guthrie Research Institute, Sayre, PA 18840; and [‡]Department of Human Genetics, Yale University, 333 Cedar Street, New Haven, CT 06510

Contributed by Sherman M. Weissman, October 24, 1988

ABSTRACT We have previously described a set of human T-lymphocyte-specific cDNA clones isolated by a modified differential screening procedure. Apparent full-length cDNAs containing the sequences of 14 of the 16 initial isolates were sequenced and were found to represent five different species of mRNA; three of the five species were identical to previously reported cDNA sequences of preproenkephalin, T-cell-replacing factor, and a serine esterase, respectively. The other two species, 4-1BB and L2G25B, were inducible sequences found in mRNA from both a cytolytic T-lymphocyte and a helper T-lymphocyte clone and were not previously described in T-cell mRNA; these mRNA sequences encode peptides of 256 and 92 amino acids, respectively. Both peptides contain putative leader sequences. The protein encoded by 4-1BB also has a potential membrane anchor segment and other features also seen in known receptor proteins.

Most T-cell factors have been classically identified by recognizing biologic activities in assays, purifying the protein, and then isolating the corresponding genes using bioactivities and protein information. An alternative approach is to isolate putative T-cell genes based upon specific expression and then to demonstrate the function of the unknown molecule (1-3). This laboratory has recently cloned a series of T-cell subset-specific cDNAs from cloned helper T-lymphocyte (HTL) L2 cells and cloned cytolytic T-lymphocyte (CTL) L3 cells by employing a modified differential screening procedure (4). Subsequently, we isolated and sequenced apparent full-length cDNAs corresponding to 14 of the original clones representing five different species of mRNA. Here we summarize our studies and present the sequences of two cDNAs that do not correspond to previously reported T-cell mRNA species.[§]

MATERIALS AND METHODS

Cells. We are grateful to F. Fitch of the University of Chicago for supplying us with the T-cell clones, L2 and L3 cells (5). Cloned murine CTL L3 cells are Thy-1.2⁺, Lyt-2⁺, LFA-1⁺, L3T4⁺, and H-2L^d-reactive. Cloned murine HTL L2 cells are Thy-1.2⁺, LFA-1⁺, Lyt-2⁺, L3T4⁺, and Mls^{a/d}-reactive.

Before the isolation of RNA L2 cells were stimulated with Con A (10 µg/ml) for 14 hr at a cell concentration of 10⁶-10⁷ cells per ml. L3 cells were stimulated with Con A (2 µg/ml) for 14 hr at a cell concentration of 2.5 × 10⁶ cells per ml. Mouse thymoma EL-4 cells (6) were stimulated with phorbol 12-myristate 13-acetate (10 ng/ml) at a cell concentration of 1.0 × 10⁶ cells per ml for 20 hr; stimulation was monitored by interleukin-2 assay (7). RNA was isolated from the unstimulated B-cell lymphoma K46 (8) and rat natural killer cell large granular lymphocytes (9).

Isolation of T-Cell-Specific cDNA Clones. We have previously isolated a group of cDNAs that are specific for T cells in contrast with B cells, employing both positive and negative differential screening and RNA blot analysis of various lymphoid cells. The T-cell-specific cDNAs were further studied to determine whether they were specific for cloned HTL L2 cells or cloned CTL L3 cells. This allowed us to isolate T-cell-subset transcripts that exist at a low level (4). The two cDNA sequences, designated as L2G25B and 4-1BB, were isolated in this fashion and were referred to as L2G25#4 and L3G29#4, respectively, in an earlier report (4).

RNA and DNA Blot Hybridization. Total cytoplasmic RNA or poly(A)⁺ RNA was fractionated on 1.2% agarose-formaldehyde gels and transferred to GeneScreenPlus (NEN). Gel-purified cDNA inserts were ³²P-labeled by nick-translation and used as probes. When a Northern (RNA) blot of GeneScreenPlus was used multiple times for hybridization, the previous probe was removed by placing the membrane in 10 mM Tris-HCl (pH 7.0)/0.2% sodium dodecyl sulfate (SDS) at 85°C for 1 hr.

High-molecular weight DNA of mouse spleens was prepared as described (10). Restriction endonuclease digests of DNA were electrophoresed in 0.8% agarose gel at 4°C. The DNA was denatured and transferred to GeneScreenPlus as described by Southern (11). The blot was hybridized with ³²P-labeled cDNA.

Screening of cDNA Library and DNA Sequencing. L2 and L3 cDNA libraries that had previously been prepared (4) were rescreened with cDNA inserts of each of the 14 T-cell-specific cDNA fragments. Typically 10 positive clones were chosen for each species, and the sizes of cDNA inserts were determined. The longest cDNA inserts were employed for nucleotide sequencing by the dideoxynucleotide chain-termination technique (12) employing Sequenase (United States Biochemical), with modifications made to accommodate 2'-deoxyadenosine 5' [α-³⁵S]thiotriphosphate (13).

Nucleotide and Protein Sequence Comparison. Full-length cDNA and predicted protein sequences were compared with the sequences in the National Institutes of Health GenBank Genetic Sequence Databank (Release 56), European Molecular Biology Laboratories (Release 16.0), and National Biomedical Research Foundation (Release 17.0). Predicted proteins were analyzed by the PEPLOT program.

RESULTS

Table 1 summarizes T-cell cDNA identified from 14-hr Con A-stimulated L2 and L3 cDNA libraries. Besides the cDNAs listed in the table, granulocyte/macrophage colony stimulating factor, interleukin 2, interleukin 3, T-cell antigen receptor α- and β-chain, and c-myc cDNAs were identified by cross-

Table 1. Summary of cDNA clones identified

Full-length cDNA	cDNA clone isolated previously*	Specificity of expression	Identification
4-1BB	L3G29#4 L3G25#4 L3G14#2	L2 and L3 L2 and L3	Unknown
L2G25B	L2G25#4 L2G95#4 L2G53#3 L2G95#3	L2 and L3	Unknown (related to cDNA PLD78)
L2S35	L2S35#3 L2PBK671	L2 only	Preproenkephalin
8-1R	L2PBK791 L2PBK642 L2PBK631	L2 and EL-4	T-cell-replacing factor
L3G10	L3G10#6 L3G18#3	L3 only	HF gene (serine esterase)
ND†	L3G7#1	L3 and EL-4	Unknown
ND†	L3G26#1	L3 and EL-4	Unknown

*cDNA clones were isolated independently and described as separate clones in the previous report (4).

†Full-length version of the two clones was not isolated.

hybridization of T-cell-enriched cDNAs with the corresponding full-length cDNA provided by other laboratories.

Among the 16 unidentified T-cell cDNA clones isolated by this screen, two clones represented preproenkephalin that was identical to the sequence reported by Zurawski *et al.* (14), three clones were T-cell-replacing factor (15), and two clones represented a T-cell serine esterase gene (16).

In the previous studies (4) we had also isolated an additional 13 L3 cDNAs whose specificity could not be assigned by RNA blot analysis because of the weak signals. One of them (L3-1) was 64% homologous to a T-cell serine esterase (17). The sequence has been reported as a new member of the

T-cell serine esterase family (18). The other 12 clones have not yet been analyzed.

Four species were from different regions of a cDNA referred to as L2G25B (800 base pairs). L2G25B was homologous to a human cDNA PLD78 (19) of unknown function. Three species (L3G29#4, L3G25#4, and L3G14#2) were from different regions of 4-1BB (2400 base pairs). We have found no reports of sequences homologous to 4-1BB. L3G7#1 and L3G26#1 were not characterized vigorously because they were represented at low levels in T-cell mRNA and were not inducible.

Fig. 1A shows the nucleotide and deduced amino acid sequence of the longest open reading frame of L2G25B. The open reading frame codes for 92 amino acids including a putative signal sequence. The deduced sequence of the first 23 amino acid residues has characteristics of the signal peptide of secretory proteins (23, 39). Therefore the mature protein is composed of 69 amino acids with a M_r of 7880. No potential N-glycosylation sites or transmembrane-like domain exists. The 3'-untranslated region has repeated occurrences of the sequence AUUUA (40) characteristic of lymphokines and other rapidly degraded mRNA.

An optimum alignment between L2G25B and PLD78 is shown in Fig. 1B. The identity of amino acids between the two proteins was ≈80%. The evolutionary conservation of these molecules from mouse to human may indicate that they play an important role in T-cell function. Recently a family of lymphokine genes termed RANTES (regulated upon activation, normal T expressed, and presumably secreted) has been defined to include PLD78 (20). After the present manuscript was prepared, a report appeared describing a macrophage inflammatory protein (21); the predicted amino acid sequence of this monokine is identical to that of L2G25B.

The nucleotide sequence of three overlapping cDNA clones represented by 4-1BB was determined (Fig. 2). The nucleotide sequence of 4-1BB revealed a single long open reading frame, beginning with the ATG codon at nucleotide

A		-47	TTTTCTG TTCTGCTGAC AAGCTCACCC TCTGTACCT GCTCAACATC	-1
1	ATG AAG GTC TCC ACC ACT GCC CTT GCT GTT CTT CTC TGT ACC ATG ACA CTC TGC AAC CAA	60		
1	Met Lys Val Ser Thr Thr Ala Leu Ala Val Leu Leu Cys Thr Met Thr Leu Cys Asn Gln	20		
61	GTC TTC TCA GCG CCA TAT GGA GCT GAC ACC CCG ACT GCC TGC TGC TTC TCC TAC AGC CGG	120		
21	Val Phe Ser Ala Pro Tyr Gly Ala Asp Thr Pro Thr Ala Cys Cys Phe Ser Tyr Ser Arg	40		
121	AAG ATT CCA CGC CAA TTC ATC GTT GAC TAT TTT GAA ACC AGC AGC CTT TGC TCC CAG CCA	180		
41	Lys Ile Pro Arg Gln Phe Ile Val Asp Tyr Phe Glu Thr Ser Ser Leu Cys Ser Gln Pro	60		
181	GGT GTC ATT TTC CTG ACT AAG AGA AAC CGG CAG ATC TGC GCT GAC TCC AAA GAG ACC TGG	240		
61	Gly Val Ile Phe Leu Thr Thr Lys Arg Asn Arg Gln Ile Cys Ala Asp Ser Lys Glu Thr Trp	80		
241	GTC CAA GAA TAC ATC ACT GAC CTG GAA CTG AAT GCC TGA GAG TCT TGC AGG CAG CGA GGA	300		
81	Val Gln Glu Tyr Ile Thr Asp Leu Glu Leu Asn Ala ---			
301	ACC CCC CAA ACC TCC ATG GGT CCC GTG TAG AGC AGG GGC TTG AGC CCC GGA ACA TTC CTC	360		
361	CCA CCT GCA TAG CTC CAT CTC CTA TAA GCT GTT TGC TGC CAA GTA GCC ATA TCG AGG GAC	420		
421	TCT TCA CTT GAA ATT TTA TTT AAT TTA ATC CTA TTG GTT TAA TAC TAT TTA ATT TTG TAA	480		
481	TTT ATT TTA TTG TCA TAC TTG TAT TTG TGA CTA TTT ATT CTG AAA GAC TTC AGG ACA CGT	540		
541	TCC TCA ACC CCC ATC TCC CTC CCA GTT GGT CAC ACT GTT TGG TGA CAG CTA TTC TAG GTA	600		
601	GAC ATG ATG ACA AAG TCA TGA ACT GAC AAA TGT ACA ATA GAT GCT TTG TTT ATA CCA GAG	660		
661	AAG TAA TAA TGC CCT TTA ACA AGT GAA AAA AAA			
B				
L2G25B (1)	Met Lys Val Ser Thr Thr Ala Leu Ala Val Leu Leu Cys Thr Met Thr Leu Cys Asn Gln			
PLD78 (1)	Met Gln Val Ser Thr Ala Ala Leu Ala Val Leu Leu Cys Thr Met Ala Leu Cys Asn Gln			
	Val Phe Ser Ala Pro Tyr Gly Ala Asp Thr Pro Thr Ala Cys Cys Phe Ser Tyr - Ser			
	- Phe Ser Ala Ser Leu Ala Ala Asp Thr Pro Thr Ala Cys Cys Phe Ser Tyr Thr Ser			
	Arg Lys Ile Pro Arg Gln - Phe Ile Val Asp Tyr Phe Glu Thr Ser Ser Leu Cys Ser			
	Arg Gln Ile Pro - Gln Asn Phe Ile Ala Asp Tyr Phe Glu Thr Ser Ser Gln Cys Ser			
	Gln Pro Gly Val Ile Phe Leu Thr Lys Arg Asn Arg Gln Ile Cys Ala Asp - Ser Lys			
	Lys Pro Gly Val Ile Phe Leu Thr Lys Arg Ser Arg Gln Val Cys Ala Asp Pro Ser -			
	Glu Thr Trp Val Gln Glu Tyr Ile Thr Asp Leu Glu Leu Asn Ala (stop) (92)			
	Glu Glu Trp Val Gln Lys Tyr Val Ser Asp Leu Glu Leu Ser Ala (stop) (92)			

FIG. 1. (A) Nucleotide sequence of L2G25B and the deduced amino acid sequence. Nucleotide sequence of the message strand is numbered in the 5' and 3' direction. The 5' noncoding sequence is indicated by negative numbers. Nucleotide residue 1 is the first nucleotide of the ATG initiation codon. The predicted amino acid sequence is shown below. Potential signal peptide is underlined, consensus polyadenylation signal is boxed, and stop codon is indicated by ---. (B) Optimum alignment between L2G25B and PLD78-deduced amino acids. Homology search revealed that the amino acid sequence of L2G25B showed an extensive homology with a reported human sequence, PLD78 (15) of unknown function, whose expression is inducible by phorbol 12-myristate 13-acetate in human tonsillar lymphocytes. The identity of amino acids between the two proteins was ≈80%. *, Identical amino acids in these proteins; +, chemically similar amino acids found in both sequences; Δ, a potential cleavage site of the signal peptide.

-145	CATGAATGCG TGAGTGGATA AACAGCACGG GATATCTCTG TCTAAGGAA TATTACTACA CCAGGAAAG	
-140	CACACATTCC ACAACAGGAA AGGAGCCTGT CACAGAAAAC CACAGTGTCC TGTGCTGTG ACATTTCGCC	
-70	1 ATG GGA AAC AAC TGT TAC AAC GTG GTG GTC ATT GTG CTG CTG CTA GTG GGC TGT GAG AAG	60
1	1 Met Gly Asn Asn Cys Tyr Asn Val Val Val Ile Val Leu Leu Leu Val Gly Cys Glu Lys	20
61	GTG GGA GCC CTG CAG AAC TCC TGT GAT AAC TGT CAG CCT GGT ACT TTC TGC AGA AAA TAC	120
21	Val Gly Ala Val Gln Asn Ser Cys Asp Asn Cys Gln Pro Gly Thr Phe Cys Arg Lys Tyr	40
121	AAT CCA GTC TGC AAG AGC TGC CCT CCA AGT ACC TTC TCC AGC ATA GGT GGA CAG CCG AAC	180
41	Asn Pro Val Cys Lys Ser Cys Pro Pro Ser Thr Phe Ser Ser Ile Gly Gly Gln Pro Asn	60
181	TGT AAC ATC TGC AGA GTG TGT GCA GGC TAT TTC AGG TTC AAG AAG TTT TGC TCC TCT ACC	240
61	Cys Asn Ile Cys Arg Val Cys Ala Gly Tyr Phe Arg Phe Lys Lys Phe Cys Ser Ser Thr	80
241	CAC AAC CGG GAG TGT GAG TGC ATT GAA GGA TTC CAT TGC TTG GGC CCA CAG TGC ACC AGA	300
81	His Asn Ala Glu Cys Glu Cys Ile Glu Gly Phe His Cys Leu Gly Pro Gln Cys Thr Arg	100
301	TGT GAA AAG GAC TGC AGG CCT GGC CAG GAG CTA ACG AAG CAG GGT TGC AAA ACC TGT AGC	360
101	Cys Glu Lys Asp Cys Arg Pro Gly Gln Glu Leu Thr Lys Lys Gly Cys Lys Thr Cys Ser	120
361	TTG GGA ACA TTT AAT GAC CAG AAC GGT ACT GGC GTC TGT CGA CCC TGG ACG AAC TGC TCT	420
121	Leu Gly Thr Phe Asn Asp Gln Asn Gly Thr Gly Val Cys Arg Pro Trp Thr Asn Cys Ser	140
421	CTA GAC GGA AGG TCT GTG CTT AAG ACC GGC ACC ACG GAG AAG GAC GTG GTG TGT GGA CCC	480
141	Leu Asp Gly Arg Ser Val Leu Lys Thr Gly Thr Thr Glu Lys Asp Val Val Cys Gly Pro	160
481	CCT GTG GTG AGC TTC TCT CCC AGT ACC ACC ATT TCT GTG ACT CCA GAG GGA GGA CCA GGA	540
161	Pro Val Val Ser Phe Ser Pro Ser Thr Thr Ile Ser Val Thr Pro Glu Gly Gly Pro Gly	180
541	GGG CAC TCC TTG CAG GTC CTT ACC TTG TTC CTG GCG CTG ACA TCG GCT TTG CTG CTG GCC	600
181	Gly His Ser Leu Gln Val Leu Thr Leu Phe Leu Ala Leu Thr Ser Ala Leu Leu Leu Ala	200
601	CTG ATC TTC ATT ACT CTC CTG TTC TCT GTG CTC AAA TGG ATC AGG AAA AAA TTC CCC CAC	660
201	Leu Ile Phe Ile Thr Leu Leu Phe Ser Val Leu Lys Trp Ile Arg Lys Lys Phe Pro His	220
661	ATA TTC AAG CAA CCA TTT AAG AAG ACC ACT GGA GCA GCT CAA GAG GAA GAT GCT TGT AGC	720
221	Ile Phe Lys Gln Pro Phe Lys Lys Thr Thr Gly Ala Ala Gln Glu Glu Asp Ala Cys Ser	240
721	TGC CGA TGT CCA CAG GAA GAA GAA GGA GGA GGA GGC TAT GAG CTG TGA TGTACTATC	780
241	Cys Arg Cys Pro Gln Glu Glu Glu Gly Gly Gly Gly Tyr Glu Leu ---	
781	CTAGGAGATG TGTGGGCGGA AACCGAGAAG CACTAGGACC CCACCATCTT GTGGAACAGC ACAAGCAACC	850
851	CCACCACCCCT GTTCTTACAC ATCATCTAG ATGATGTGTG GCGCGCACCC TCATCCAAGT CTCTCTAAC	920
921	GCTAACATAT TTGCTTTTAC CTTTTTTTAA TCTTTTTTAA AATTAAATTT TTATGTGTGT GAGTGTTTG	990
991	CCTGGCTGTA TGCACACGGT TGTGTGTGTG TGTGTGTGAC ACTCCTGATG CCGGAGGAGG TCAGAAGAGA	1060
1061	AAGGCTGGT TCCATAAGAA CTGGAATTTT GATGGCTGTG GAGCGGnnn GATAGGTGGG GACGGAGACC	1130
1131	TGTCTTCTTA TTTAAACGTG ACTGTATGAT AAAAAAAAAA TGATATTTCC GAAATGTAG AGATTGCTCT	1200
1201	GACACCCCTC TAGTTAATGA TCTAAGAGGA ATTGTTGATA CGTAGTATAC TGATATGTG TATGTAATG	1270
1271	TATATGTATA TATAAGACTC TTTTACTGTC AAAGTCAACC TAGAGTGTCT GGTACCAGG TCAATTTAT	1340
1341	TGGACATTTT ACGTACACACA CACACACACA CACACACACA CAGTTTATA CTACGTACTGT TATCGGTAT	1410
1411	TCTACGTCAT ATAATGGGAT AGGCTAAAG GAAACCAAG AGTGAGTAT ATTTATTTGGA GTGACAGA	1480
1481	CTACCCCTTC TGGGTACGTA GGGACAGACC TCCTTGGGAC TGTCTAAAC TCCTCTAGA AGTCTGCTA	1550
1551	AGTTCGCCGA CGAACAGGAC AGAGGACACA CAGTCCGAAA AGTTATTTT CCGGCAAAATC CTTTCCCTGT	1620
1621	TTGCTGACAC TCACCCCTCT GTGACACCTT GAGTGTCTC CTTCGCCGGG AAGTCCAGT GGTACCCCTC	1690
1691	TGTAGGGGGG GGGAGACAGA GCGCGGGGGG AGCTACGACA ATCGACTCAC AGGGCCGCCG GGGTCTGCA	1760
1761	AATGAAACTT TTTTAACTTC ACAAGTTTTC TCCGGGCTCG GCGGACCTAT GCGGTGTGAT CTTATTACCT	1830
1831	TATCTCGGCG CCAAGATAAA ACAACCAAAA GCCTTGACTC CGGTACTAAT TCTCCCTGCC GCGCCCGCTA	1900
1901	AGCATACGCG GCGGATCTCC ACTTTAAGAA CCGTGGCGGG TTCTGCTGG TCTCGCTTC GTAACCGGTT	1970
1971	CTTACAAAGG TAATTAGTTC TTGCTTTTCA CCTCAAGCT TCTGCTAGTC TATGGCAGCA TCAAGCGCTG	2040
2041	TATTGTCTAC GCGTCAGCCG TACGCCCGCG CAATAAGGCT ACTGGCGCCG CCGTCAAGG CCTTTTGGTT	2110
2111	TCAGAAACCC AAGGCCCGCC TCATACCAAC GTTTCGACTT TGATTCTTGC CGGTACGTGG TGTGGGGTGC	2180
2181	CTTAGCTCTT TCTCGATAGT TAG AC	

FIG. 2. Nucleotide sequence and the deduced amino acid sequence of 4-1BB. The nucleotides of the message strand are numbered in the 5' and 3' direction, and numbers are shown on both sides of the sequence. Nucleotide residue 1 is the adenine of the initiation codon ATG, and the nucleotides on the 5' side of residue 1 are indicated by negative numbers. The predicted amino acid sequence is shown below the nucleotide sequence. Putative signal peptide and the potential asparagine-linked glycosylation sites are underlined; the potential transmembrane domain is double underlined. The potential polyadenylation signal is boxed. N, ambiguous nucleotide sequences in the 3' untranslated region; ---, stop codon; •, cysteine residues.

residues 1–3. This reading frame codes for a polypeptide of 256 amino acids with a M_r of 27,587. The assigned ATG is preceded by an in-frame termination codon TGA (nucleotide residues –12 to –9). The sequence flanking the assigned ATG (nucleotide residues –5 to 4) contains eight of nine residues identical to the consensus sequence (CCRC-CATGG, where R represents guanosine or adenine) described by Kozak (22). The codon specifying the carboxyl-terminal leucine is followed by the translational termination codon TGA (nucleotide residues 769–771). Clone 4-1BB contains an unusually long 3'-untranslated sequence, which did not extend as far as the poly(A)⁺ tail. A potential polyadenylation signal is present at nucleotides 1158–1163 (Fig. 2, boxed). This signal may sometimes be functional because this gene produces at least two different sizes of mRNA. Fig. 3 *a* and *b* shows RNA blot analysis of Con A-stimulated L3 RNA. When the blot was hybridized to an L3G25#4 probe that contained sequences to the 3' side of the polyadenylation signal (nucleotides 1284–1557), the probe detected one RNA species of ≈2.4 kilobases (kb). When the

same blot was hybridized to an L3G14#2 probe that contained sequences to the 5' side of the polyadenylation signal (nucleotides 661–855) the probe detected two mRNA species of ≈1.5 kb and 2.4 kb. Only one copy of the sequence ATTTA that is often present in multiple copies in the 3'-untranslated regions of unstable mRNA was present in 4-1BB, and further studies of the stability of its mRNA are needed. Northern blot analysis has shown that 4-1BB mRNA is inducible in both L2 and L3 cells and is present in several CTL clones and hybridomas but is absent in the EL-4 T-cell line (B.S.K., unpublished results).

The deduced sequence of the first 23 amino acids of 4-1BB cDNA has characteristics of the signal peptide of secretory and membrane-associated proteins (23) and fits the –1, –3 rule (24). We putatively assigned the first 23 amino acids as a signal peptide, although the presence of lysine at –4 and glutamic acid at –5 are somewhat unusual. Thus, the protein backbone of processed 4-1BB protein would be composed of 233 amino acids with a M_r of 25,167. Two potential asparagine-linked glycosylation signals (25, 26) are located at

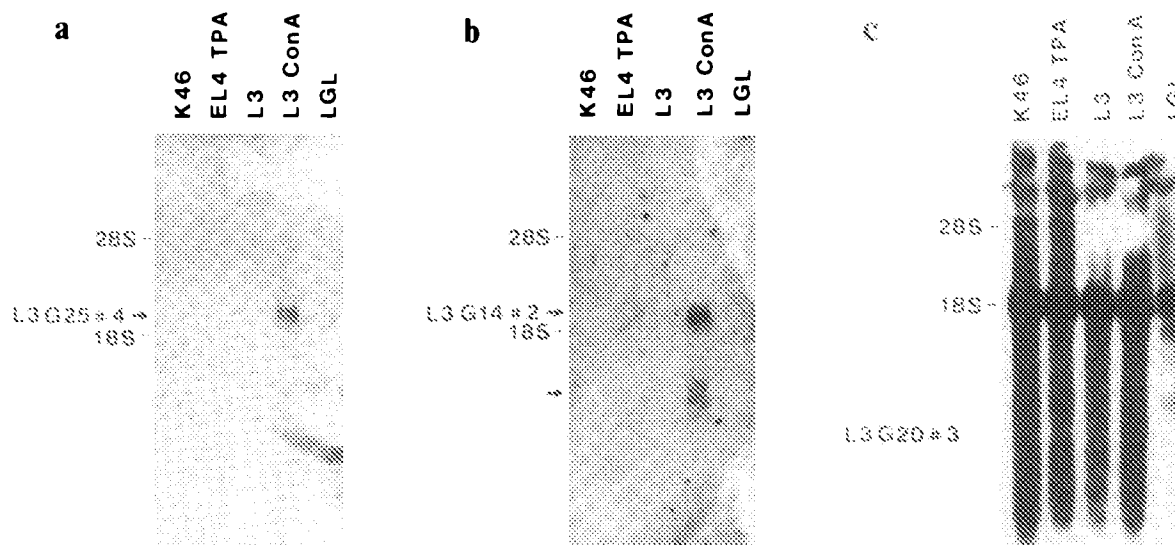


FIG. 3. Expression of two different sizes of 4-1BB mRNA. Ten micrograms of poly(A)⁺ mRNA from mouse B-cell line (K46), phorbol 12-myristate 13-acetate-stimulated EL-4 (EL-4 TPA), and rat natural killer cell line (LGL) and 10 μ g of total RNA from unstimulated L3 (L3) and Con A-stimulated L3 (L3 ConA) were fractionated on a 1.4% formaldehyde-agarose gel, transferred to GeneScreenPlus, and hybridized to ³²P-labeled L3G25#4 (a), L3G14#2 (b), and L3G20#3 (c) sequentially. L3G25#4 and L3G14#2 represent cDNA fragments of the 3' side and 5' side to boxed AATAAA sequence, respectively. L3G20#3 is an anonymous cDNA from L3 cDNA library and is used to show that each lane of the blot contains a similar amount of RNA. Positions of 28S and 18S rRNA markers are each indicated; arrows indicate the specific hybridization signals.

amino acid positions 128 and 138 as underlined in Fig. 2. The predicted 4-1BB protein contained an unusually large number of cysteines. There are 23 cysteine residues in the putative mature protein as marked with dots in Fig. 2. These residues are arranged with a spacing reminiscent of that seen in several groups of proteins including zinc finger DNA-binding proteins (27), epidermal growth factor receptor (28), the *Drosophila* Notch locus product (29), and certain translation factors (30, 31). Amino acids flanking certain of the cysteine groups also resemble sequences found in other proteins. For example, with a replacement of leucine by isoleucine, there is an exact match to seven of eight amino acids of a putative zinc finger structure in the yeast eIF-2 β protein (31). These residues may represent metal binding sites, but, as discussed by others (32), their role in the functions of a protein cannot be uniquely predicted.

Following the cysteine-rich regions of the protein is a stretch of amino acids (residues 140–185) in which almost 30% of the amino acids are serine or threonine; these are potential sites for O-linked glycosylation and are reminiscent of those seen, for example, in the low density lipoprotein receptor (33). There is a stretch of 26 amino acids that constitutes a hydrophobic domain toward the carboxyl terminus of the protein (amino acids 186–211) (Fig. 4), followed by a sequence containing several basic residues. This region may serve as a membrane-spanning anchor domain.

Finally, the relatively short carboxyl-terminal putative cytoplasmic domain contains both two short runs of three and four acidic residues, respectively, and a sequence of five glycines followed by a tyrosine. The short stretches of acidic residues are also found in the cytoplasmic domain of the low

density lipoprotein and atrial natriuretic peptide receptors, for example (33, 35).

Southern Blot Analysis. As shown in Fig. 5, a fragment of L2G25B and of 4-1BB cDNA each detects a single-restriction fragment of \approx 15 kb and 18 kb in C57BL/6 and BALB/c DNA, respectively. The data indicated that the genes encoding the two molecules exist as a single copy in C57BL/6 and BALB/c mice.

DISCUSSION

We and others have developed protocols for differential screening of cDNA libraries by which one can detect a broad representation of the mRNA differentially expressed in two cell types (1–4). The method was applied to the systematic analysis of HTL and CTL gene expression and offered an alternative approach to classical protein purification for identifying effector molecules and genes. The genes for the T-cell antigen receptor, certain T-cell-specific effectors, and the probable gene mutated in an X chromosome-linked immunodeficiency syndrome of mice (36) were cloned and characterized in this fashion, and the set of clones described in this and a previous report were obtained before most of the individual species represented had been isolated by other methods.

A number of lymphokine clones, including interleukins 2, 3, and 5 and granulocyte/macrophage colony-stimulating factor as well as other T-cell-specific clones such as the serine esterases and T-cell receptors, were selected in the relatively small libraries (total of 18,000 L2 clones and 10,000 L3 clones) whose screening we have reported here and elsewhere. Probably the bulk of the most abundant T-cell-specific

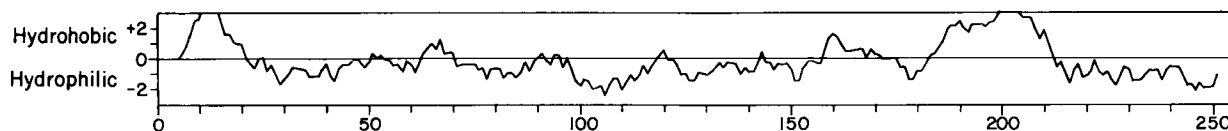


FIG. 4. Hydropathicity profile of the predicted amino acid sequence of 4-1BB. Local hydropathicity values calculated by the method of Kyte and Doolittle (34) were plotted versus amino acid residues. Positive values indicate hydrophobic regions, and negative values indicate hydrophilic regions.

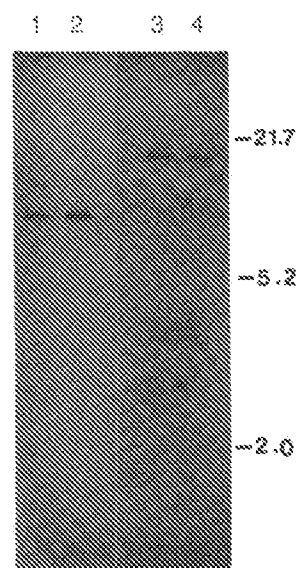


FIG. 5. Southern blot analysis of mouse genomic DNA. Genomic DNA from C57BL/6 (lanes 1 and 3) and BALB/c (lanes 2 and 4) was digested with *Eco*RI restriction enzyme, fractionated on a 0.8% agarose gel, transferred to GeneScreenPlus, and hybridized to ³²P-labeled L2G25B (lanes 1 and 2) and 4-1BB (lanes 3 and 4).

sequences were detected. Nevertheless, other potent biologic effectors may well have been missed. For example, L3 secretes large amounts of interferon γ , but, for unknown reasons, we did not recover a cDNA clone for this molecule. No clone for any member of the RANTES family other than L2G25B was detected, even though mRNA for such genes may be among the most abundant mRNA species produced in T cells. Similarly, clones for other T-cell activation antigens and T-cell-specific membrane channel-forming proteins were not detected.

Clones of the cDNA of 4-1BB were obtained from both the L2 and L3 cells. The clones corresponding to L2G25B were all isolated from L2 cells, and the possibility remains that the mRNA sequences detected in L3 corresponded to a different member of this gene family. The two species of cDNA reported here represent the two abundant, inducible sequences detected in both the HTL and CTL clones used in the present study. The induction properties of both clones resembled those of other lymphokines in that induction could be detected within 30 min after stimulation; they could both be induced either by Con A or by antigen stimulation. Unlike interleukin 2 receptor mRNA (37), induction of both mRNA species was blocked by treatment with cyclosporine A (B.S.K., unpublished work). Because 4-1BB protein was inducible in both T-cell clones, it may play roles in the T-cell activation process and function of the activated T cells less specific to a certain T-cell subset than those of the well-described lymphokines. As discussed above, 4-1BB has a number of properties reminiscent of membrane receptors and transmembrane signaling proteins. However, in at least the instances of the *Drosophila* Notch and sevenless loci (29, 38), an apparent transmembrane protein seems capable of acting directly or indirectly *in trans* on other cells.

We thank Drs. Frank Fitch, David Lancki, and Mike Prystowsky for providing the L2 and L3 cell preparations; Dr. Pierre Henkart for large granular lymphocytes; Dr. John Farrar for EL-4 and CT-6 cells; Dr. Charles Janeway for A20.2J and K46 cells; Dr. Steve Litwin and Mrs. Mary Kiefer for editing the manuscript; Drs. Asifa Haq, Daniel Kestler, Mrs. Alex Liddell, and Mr. Mark Wakulchik for technical help; and Mrs. Helen Kelley, Mrs. Mary Kiefer, and Mrs. Elaine Wall for typing the manuscript. This work was supported in part by Grants AI23058 and AI28175 from the National Institutes of Health and the Feasibility Grant program of the American Diabetes Association.

1. Hedrick, S., Nielsen, E., Kavalier, J., Cohen, D. & Davis, M. (1984) *Nature (London)* **308**, 153–158.
2. Saito, H., Krantz, D., Takagaki, Y., Hayday, A., Eisen, H. & Tonegawa, S. (1984) *Nature (London)* **312**, 36–40.
3. Burd, P., Freeman, G., Wilson, S., Berman, M., Dekruyff, R., Billings, P. & Dorf, M. (1987) *J. Immunol.* **139**, 3126–3131.
4. Kwon, B., Kim, G., Prystowsky, M., Lancki, D., Sabath, D., Pan, J. & Weissman, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2896–2900.
5. Glasebrook, A. & Fitch, F. (1980) *J. Exp. Med.* **151**, 876–895.
6. Farrar, J., Fuller-Farrar, J., Simon, P., Hilfiker, M., Stadler, B. & Farrar, W. (1980) *J. Immunol.* **125**, 2555–2558.
7. Gillis, S., Ferm, M., Ou, W. & Smith, K. (1978) *J. Immunol.* **120**, 2027–2032.
8. Kim, K., Kanellopoulos-Langevin, C., Merwin, R., Sach, D. & Asofsky, R. (1979) *J. Immunol.* **122**, 549–554.
9. Henkart, P., Millards, P., Reynolds, C. & Henkart, M. (1984) *J. Exp. Med.* **160**, 75–93.
10. Grass-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32–38.
11. Southern, E. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Biggin, M., Gibson, T. & Hung, G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
14. Zurawski, G., Benedik, M., Kamb, B. J., Abrams, J. S., Zurawski, S. M. & Lee, F. D. (1986) *Science* **232**, 772–775.
15. Kinachi, T., Harada, N., Severinson, E., Tanabe, T., Sideras, P., Konishi, M., Azuma, C., Tominaga, A., Bergstedt-Lindquist, S., Takahashi, M., Matsuda, F., Yaoita, Y., Takatsu, K. & Honjo, T. (1986) *Nature (London)* **325**, 70–73.
16. Gershenfeld, H. K. & Weissman, I. L. (1986) *Science* **232**, 854–858.
17. Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. M. & Bleackley, R. C. (1986) *Science* **232**, 858–861.
18. Kwon, B., Kestler, D., Lee, E., Wakulchik, M. & Young, J. (1988) *J. Exp. Med.* **168**, 1839–1854.
19. Obaru, K., Fukuda, M., Maeda, S. & Shimada, K. (1986) *J. Biochem.* **99**, 885–894.
20. Schall, T., Jongstra, J., Dyer, B., Jorgensen, J., Clayberger, C., Davis, M. & Krensky, A. (1988) *J. Immunol.* **141**, 1018–1025.
21. Davatellis, G., Tekamp-Olson, P., Wolpe, S., Hermesen, K., Lyedke, C., Gallegos, C., Coit, D., Merryweather, J. & Cerami, A. (1988) *J. Exp. Med.* **167**, 1939–1944.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
23. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 852–862.
24. Von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
25. Marshall, R. D. (1974) *Biochem. Soc. Symp.* **40**, 17–26.
26. Bause, E. (1983) *Biochem. J.* **209**, 331–336.
27. Evans, R. & Hollenberg, S. (1988) *Cell* **52**, 1–3.
28. Ullrich, A., Coussens, L., Hayflick, J., Dull, T., Gray, A., Tam, A., Lee, J., Yarden, Y., Liebermann, T., Schlessinger, J., Downward, J., Mayes, E., Whittle, N., Waterfield, M. & Seeborg, P. (1984) *Nature (London)* **309**, 418–425.
29. Wharton, K., Johansen, K., Xu, T. & Artavanis-Tsakonas, S. (1985) *Cell* **43**, 567–581.
30. Pathak, V., Nielsen, P., Trachsel, H. & Hershey, J. (1988) *Cell* **54**, 633–639.
31. Donahue, T., Cigan, A., Pahich, E. & Valavicius, B. (1988) *Cell* **54**, 621–632.
32. Frankel, A. & Pabo, C. (1988) *Cell* **53**, 675.
33. Yamamoto, T., Davis, C., Brown, M., Schneider, W., Casey, M., Goldstein, J. & Russell, P. (1984) *Cell* **39**, 27–38.
34. Kyte, J. & Doolittle, R. (1982) *J. Mol. Biol.* **157**, 105–132.
35. Fuller, F., Porter, J., Arfsten, A., Miller, J., Schilling, J., Scarborough, R., Lewicki, J. & Schenk, D. (1988) *J. Biol. Chem.* **263**, 9395–9401.
36. Cohen, D., Steinberg, A., Paul, W. & Davis, M. (1985) *Nature (London)* **314**, 372–374.
37. Kronke, M., Leonard, W., Depper, J., Arya, S., Wong-Staal, F., Gallo, R., Waldman, T. & Green, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5214–5218.
38. Hafen, E., Basler, K., Edstroem, J.-E. & Rubin, G. (1987) *Science* **236**, 55–63.
39. Steiner, D., Quinn, P., Chan, S., Marsh, J. & Tager, H. (1980) *Ann. N.Y. Acad. Sci.* **34**, 1–16.
40. Shaw, G. & Kamen, R. (1986) *Cell* **46**, 659–667.